

# Variance Component Estimation Using the Additive, Dominance, and Additive $\times$ Additive Model When Genotypes Vary across Environments

Jixiang Wu, Johnnie N. Jenkins,\* Jack C. McCarty, and Dongfeng Wu

## ABSTRACT

In addition to additive (A) and dominant (D) genetic effects, the  $A \times A$  interaction (or  $A \times A$  epistatic) effects that control many quantitative traits are important for genetic and breeding studies. To estimate these genetic variance components, including genotype  $\times$  environment ( $G \times E$ ) interaction, one usually expects to have data from at least two generations (i.e.,  $F_1$  and  $F_2$ ) and parents with the same entries in all environments. Practical difficulties may arise in implementing such a design. In this study, we performed Monte Carlo simulations to compare the estimated variance components between four partial and two complete genetic designs (GDs) using the mixed linear model approach. Our definition for GD is different from the traditional definitions of genetic mating designs. Simulation results showed that the estimated genetic variance components for A,  $A \times E$ ,  $A \times A$  epistatic, and  $A \times A \times E$  effects were unbiased for the six designs. Among four partial designs, two provided the comparable results for D and  $D \times E$  effects compared with the complete GDs, but with slightly larger mean square errors (MSEs), indicating that some partial GDs can be used when the genetic resources are limited.

ESTIMATION OF GENETIC VARIANCE components is an important issue in the field of quantitative genetics. A genetic mating design is often required to reach this goal. The widely used genetic mating designs for estimating A and D variance components include nested mating design (North Carolina I), factorial mating design (North Carolina II), and various types of diallel mating designs (Griffing's mating designs). Readers interested in these genetic mating designs may refer to Chapters 18 and 19 in Lynch and Walsh's book (1998). If a full diallel mating design is used, maternal and/or paternal effects may be detected. Usually, the ANOVA approach is used to estimate these genetic variance components. However, there are some limitations when using the ANOVA approach: (i) the mating design should be balanced (with no missing crosses), (ii) it requires the  $F_1$  hybrids, and (iii) the genetic model is not extendable. Thus, the ANOVA approach has limitations in selecting genetic models, generations, or data sets.

In addition to A and D effects, the epistatic (gene  $\times$  gene interaction at different gene loci) effects are important effects and influence many quantitative traits in nature. Several quantitative trait locus mapping studies

have shown epistatic effects (Cao et al., 2001; Doebley et al., 1995; Eshed and Zamir, 1996; Lark et al., 1995; Lee et al., 2001; Li et al., 1997; Liao et al., 2001; Wu, 2003). Few reports for the epistatic effects in the traditional quantitative genetic study could be found in literature (Goodnight, 1988; Cheverud and Routman, 1996; Xu and Zhu, 1999; Edwards and Lamkey, 2002; McCarty et al., 2004; Saha et al., 2004). Among the many possible epistatic effects, the  $A \times A$  (or  $A \times A$  epistatic) effects are one of the most important one for breeding studies. A genetic model containing A effects, D effects, and  $A \times A$  epistatic effects is called the ADAA model (Cockerham, 1980). This model can also be extended to include  $G \times E$  interaction effects (Zhu, 1994). For the ADAA genetic model, data only containing  $F_1$ s and parental lines may not be sufficient. On the basis of Zhu's study (Zhu, 1994, 1998), at least two generations such as  $F_1$ s and  $F_2$ s and parents are required if the ADAA model is used. Since  $G \times E$  interactions can have a significant influence on a quantitative trait, it is desirable that such traits be measured under multiple environments. Ideally, the genotypes under different environments should be balanced (or treated under a complete GD). In reality, it may be difficult to evaluate as a complete GD due to practical reasons. Even so, researchers want to get maximum genetic information from these experiments with unbalanced designs. The ANOVA or GLM methods are inappropriate for the analysis. Mixed linear model approaches offer the flexibility of analyzing various types of unbalanced data (Zuo et al., 2000, 2001; McCarty et al., 2004); however, the feasibility of using partial GDs for efficient estimation of genetic variance components for the ADAA genetic model remains unknown and warrants further investigation.

In this study, we provide several different GDs (complete and partial) and evaluate the estimation for variance components via simulation. This study provides important alternatives to guide breeders or experimenters in utilizing genetic experiments when complete GDs under multiple environments are difficult to implement.

## GENETIC MODELS AND METHODOLOGY

### Genetic Model

The genetic model including A, D,  $A \times A$  effects, and their corresponding  $G \times E$  interaction effects (ADAA model) was employed for the data analysis (Cockerham, 1980; Zhu, 1994).

The mixed linear models were as follows:

**Abbreviations:**  $\sigma^2_\mu$ , variance component mean; A, additive; D, dominance or dominant; E, environment; G, genotype; GD, genetic design; MINQUE, minimum norm quadratic unbiased estimation; MSE, mean square error.

J. Wu, Dep. of Plant and Soil Sciences, J.N. Jenkins and Jack C. McCarty, Crop Science Research Laboratory, USDA-ARS, D. Wu, Dep. of Mathematics and Statistics, Mississippi State Univ., Mississippi State, MS 39762. \*Corresponding author (jnjenkins@ars.usda.gov).

Published in Crop Sci. 46:174–179 (2006).

Crop Breeding Genetics & Cytology

doi:10.2135/cropsci2005.04-0025

© Crop Science Society of America

677 S. Segoe Rd., Madison, WI 53711 USA

Parents:

$$y_{hiik(P)} = \mu + E_h + 2A_i + D_{ii} + 4AA_{ii} + 2AE_{hi} + DE_{hii} + 4AAE_{hii} + B_{k(h)} + e_{hiik} \quad [1]$$

F<sub>1</sub>:

$$y_{hiik(F_1)} = \mu + E_h + (A_i + A_j) + D_{ij} + (AA_{ii} + AA_{jj} + 2AA_{ij}) + (AE_{hi} + AE_{hj}) + DE_{hii} + (AAE_{hii} + AAE_{hjj} + 2AAE_{hij}) + B_{k(h)} + e_{hiik} \quad [2]$$

F<sub>2</sub>:

$$y_{hiik(F_2)} = \mu + E_h + (A_i + A_j) + (1/4D_{ij} + 1/4D_{jj} + 1/2D_{ij}) + (AA_{ii} + AA_{jj} + 2AA_{ij}) + (AE_{hi} + AE_{hj}) + (1/4DE_{hii} + 1/4DE_{hjj} + 1/2DE_{hij}) + (AAE_{hii} + AAE_{hjj} + 2AAE_{hij}) + B_{k(h)} + e_{hiik} \quad [3]$$

F<sub>3</sub>:

$$y_{hiik(F_3)} = \mu + E_h + (A_i + A_j) + (3/8D_{ij} + 3/8D_{jj} + 1/4D_{ij}) + (AA_{ii} + AA_{jj} + 2AA_{ij}) + (AE_{hi} + AE_{hj}) + (3/8DE_{hii} + 3/8DE_{hjj} + 1/4DE_{hij}) + (AAE_{hii} + AAE_{hjj} + 2AAE_{hij}) + B_{k(h)} + e_{hiik} \quad [4]$$

Where  $\mu$  is the population mean, a fixed effect;  $E_h$  is the environment effect, either random or fixed (fixed in this study);  $A_i$  (or  $A_j$ ) is additive effect from parent  $i$  (or  $j$ );  $D_{ii}$ ,  $D_{jj}$ , or  $D_{ij}$  is the dominance effect;  $AA_{ii}$ ,  $AA_{jj}$ , or  $AA_{ij}$  is the additive  $\times$  additive epistatic effect;  $AE_{hi}$  (or  $AE_{hj}$ ) is additive  $\times$  environment interaction effect;  $DE_{hii}$ ,  $DE_{hjj}$ , or  $DE_{hij}$  is the dominance  $\times$  environment interaction effect;  $AAE_{hii}$ ,  $AAE_{hjj}$ , or  $AAE_{hij}$  is the additive  $\times$  additive  $\times$  environment interaction effect;  $B_{k(h)}$  is the block effect; and  $e_{hiik}$  is the random error.

Equations [1] to [4] can be expressed in the form of vectors and matrices as follows.

$$\mathbf{y} = \mathbf{1}\mu + \mathbf{X}_E\mathbf{b}_E + \mathbf{U}_A\mathbf{e}_A + \mathbf{U}_D\mathbf{e}_D + \mathbf{U}_{AA}\mathbf{e}_{AA} + \mathbf{U}_{AE}\mathbf{e}_{AE} + \mathbf{U}_{DE}\mathbf{e}_{DE} + \mathbf{U}_{AAE}\mathbf{e}_{AAE} + \mathbf{U}_B\mathbf{e}_B + \mathbf{e} = \mathbf{X}\mathbf{b} + \sum_{u=1}^8 \mathbf{U}_u\mathbf{e}_u \quad [5]$$

In this model, we assume that the E effect is fixed, where  $\mu$  is the population mean;  $\mathbf{1}$  is the vector with all elements 1;  $\mathbf{e}_A$  is the vector for additive effects,  $\mathbf{e}_A \approx N(0, \sigma_A^2\mathbf{I})$ ;  $\mathbf{U}_A$  is the incidence matrix for additive effects,  $\mathbf{e}_D$  is the vector for dominance effects,  $\mathbf{e}_D \approx N(0, \sigma_D^2\mathbf{I})$ ;  $\mathbf{U}_D$  is the incidence matrix for dominance effects;  $\mathbf{e}_{AA}$  is the vector for additive  $\times$  additive effects,  $\mathbf{e}_{AA} \approx N(0, \sigma_{AA}^2\mathbf{I})$ ;  $\mathbf{U}_{AA}$  is the incidence matrix for additive  $\times$  additive effects;  $\mathbf{e}_{AE}$  is the vector for additive  $\times$  environment effects,  $\mathbf{e}_{AE} \approx N(0, \sigma_{AE}^2\mathbf{I})$ ;  $\mathbf{U}_{AE}$  is the incidence matrix for additive  $\times$  environment effects;  $\mathbf{e}_{DE}$  is the vector for dominance  $\times$  environment effects,  $\mathbf{e}_{DE} \approx N(0, \sigma_{DE}^2\mathbf{I})$ ;  $\mathbf{U}_{DE}$  is the incidence matrix for dominance  $\times$  environment effects;  $\mathbf{e}_{AAE}$  is the vector for additive  $\times$  additive  $\times$  environment effects,  $\mathbf{e}_{AAE} \approx N(0, \sigma_{AAE}^2\mathbf{I})$ ;  $\mathbf{U}_{AAE}$  is the incidence matrix for AAE effects;  $\mathbf{e}_B$  is the vector for block effects,  $\mathbf{e}_B \approx N(0, \sigma_B^2\mathbf{I})$ ;  $\mathbf{U}_B$  is the incidence matrix for block effects;  $\mathbf{e}_e$  is the vector for random errors,  $\mathbf{e}_e \approx N(0, \sigma_e^2\mathbf{I})$ .

### Variance Component Estimation

The minimum norm quadratic unbiased estimation (MINQUE) approach was proposed by Rao (1971) for esti-

**Table 1. Six genetic designs for the ADAA model.<sup>†</sup>**

Genetic design	Environment	Entries
1	1	7P + 21F <sub>1</sub> + 21F <sub>2</sub>
	2	7P + 21F <sub>1</sub> + 21F <sub>2</sub>
2	1	7P + 21F <sub>2</sub> + 21F <sub>3</sub>
	2	7P + 21F <sub>2</sub> + 21F <sub>3</sub>
3	1	7P + 21F <sub>1</sub>
	2	7P + 21F <sub>2</sub>
4	1	7P + 21F <sub>2</sub>
	2	7P + 21F <sub>3</sub>
5	1	7P + 21F <sub>1</sub>
	2	7P + 21F <sub>1</sub> + 21F <sub>2</sub>
6	1	7P + 21F <sub>2</sub>
	2	7P + 21F <sub>2</sub> + 21F <sub>3</sub>

<sup>†</sup>ADAA = a genetic model containing additive (A) effects, dominant (D) effects, and A  $\times$  A epistatic effects.

imating variance components. The variance components in the ADAA model can be estimated by solving the following MINQUE normal equations for  $u, v = 1, 2, \dots, 8$ .

$$[\text{tr}(\mathbf{U}_u^T \mathbf{Q}_\alpha \mathbf{U}_v^T \mathbf{U}_v \mathbf{Q}_\alpha \mathbf{U}_u)] [\sigma_u^2] = [\mathbf{y}^T \mathbf{Q}_\alpha \mathbf{U}_u \mathbf{U}_u^T \mathbf{Q}_\alpha \mathbf{y}] \quad [6]$$

where the trace (tr) is the sum of diagonals of a matrix, and

$$\mathbf{Q}_\alpha = \mathbf{V}_\alpha^{-1} - \mathbf{V}_\alpha^{-1} \mathbf{X} (\mathbf{X}^T \mathbf{V}_\alpha^{-1} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{V}_\alpha^{-1}, \quad [7]$$

where  $\mathbf{V}_\alpha = \sum_{u=1}^8 \alpha_u \mathbf{U}_u \mathbf{U}_u^T$  and  $\mathbf{V}_\alpha^{-1}$  is the inverse matrix of  $\mathbf{V}_\alpha$  with prior values  $\alpha_u$  in place of  $\sigma_u^2$  in  $\mathbf{V}$ . In this simulation study, we set  $\alpha_u = 1, u = 1, \dots, 8$ , so this approach is called MINQUE(1) (Zhu, 1989).

### Genetic Designs

Six GDs evaluated are listed in Table 1. Note that a partial GD in this study could be where the genotypes or generations are different across two environments and a complete GD is one where genotypes and generations used in two environments are exactly the same. Thus, our notations for the partial and complete GDs are different from the partial and complete mating GDs defined by Comstock and Robinson (1948, 1952) and Griffing (1956). Both partial and complete GDs could include the genetic materials obtained from either a complete or partial mating design. Designs 1 and 2 were called complete GDs because genotypes and generations across two environments were the same, while Designs 3 to 6 were called partial GDs because genotypes or generations were different in the two environments (Table 1).

### Simulation

For simplicity, a randomized complete block design with four replicates in each environment was used in this simulation study. Note that in this simulation study, the number of genotypes could be different between the two environments (Table 1). The designs are used based on a set of half-diallel crossed of seven parents. On the basis of the six designs in Table 1, we generated the phenotypic data with each effect vector following a normal distribution,  $e_u \approx N(0, \sigma_u^2\mathbf{I})$ . Variance components were estimated by the MINQUE(1) approach (Zhu, 1989). Each GD with three sets of input values was analyzed with 500 simulations. In addition to calculating each variance component mean ( $\bar{\sigma}_u^2$ ), the corresponding bias was calculated accordingly by  $\bar{\sigma}_u^2 - \sigma_u^2$ , where  $\sigma_u^2$  was the preset value. The MSE for each parameter was calculated by  $\text{MSE} = \text{bias}^2 + \text{var}(\bar{\sigma}_u^2)$ . All Monte Carlo simulation programs and program for actual data analysis were written in C++ language.

**Table 2. Simulations results for the ADAA model with existence of all genetic effects by 500 simulations.**

	Value	GD 1†		GD 2		GD 3		GD 4		GD 5		GD 6	
		Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
$\sigma_A^2$	20	0.98	7.55	1.19	8.63	1.74	12.36	1.96	13.17	–2.20	11.63	–2.31	12.39
$\sigma_D^2$	20	–0.29	0.67	2.82	10.09	–11.21	270.23	–20.00	868.58	0.70	1.87	–1.72	12.05
$\sigma_{AA}^2$	20	0.24	0.27	0.51	0.49	1.77	5.41	1.69	3.53	0.50	0.53	0.73	0.79
$\sigma_{AE}^2$	20	–1.37	4.47	2.17	7.69	1.45	7.33	–2.88	12.95	2.03	7.04	2.92	11.57
$\sigma_{DE}^2$	20	0.02	0.47	–3.67	16.16	12.38	288.46	42.73	2448.99	0.05	1.32	2.04	14.02
$\sigma_{AAE}^2$	20	–0.07	0.12	–0.69	0.56	–1.32	3.91	–1.53	3.13	–0.10	0.16	0.18	0.15
$\sigma_e^2$	20	–0.01	0.01	–0.04	0.01	0.04	0.01	–0.10	0.02	–0.15	0.03	–0.06	0.01

† Genetic Designs 1 through 6 (GD1–GD6) are defined in Table 1.

**Table 3. Simulations results for the ADAA model with no  $A \times A$  epistatic effects by 500 simulations.**

	Value	GD 1†		GD 2		GD 3		GD 4		GD 5		GD 6	
		Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
$\sigma_A^2$	20	0.69	1.47	1.18	2.31	1.98	5.00	1.38	2.91	1.32	2.87	2.85	9.15
$\sigma_D^2$	20	–0.38	0.73	2.81	10.04	2.65	14.26	–4.06	51.75	0.70	1.03	0.67	2.22
$\sigma_{AA}^2$	0	0.73	0.54	0.54	0.29	0.23	0.17	0.72	0.56	0.74	0.56	0.58	0.34
$\sigma_{AE}^2$	20	–0.71	0.87	–0.66	0.81	–1.06	1.69	–1.85	3.92	–0.86	1.13	–0.17	0.40
$\sigma_{DE}^2$	20	0.15	0.49	–3.75	16.74	–4.49	28.41	19.10	401.60	0.21	0.53	–0.42	2.86
$\sigma_{AAE}^2$	0	0.32	0.10	0.31	0.10	1.04	1.22	0.00	0.04	0.26	0.09	0.34	0.12
$\sigma_e^2$	20	–0.01	0.01	–0.04	0.01	0.04	0.01	–0.10	0.02	–0.15	0.03	–0.05	0.01

† Genetic Designs 1 through 6 (GD1–GD6) are defined in Table 1.

**Table 4. Simulations results for the ADAA model with no D effects by 500 simulations.**

	Value	GD 1†		GD 2		GD 3		GD 4		GD 5		GD 6	
		Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
$\sigma_A^2$	20	0.33	5.58	–0.64	6.71	–0.10	8.37	–0.49	8.33	–3.15	15.51	–2.44	12.10
$\sigma_D^2$	0	0.53	0.38	2.08	5.43	0.00	112.83	0.00	416.87	1.13	1.68	0.00	6.05
$\sigma_{AA}^2$	20	–0.36	0.34	–0.09	0.22	1.47	4.07	0.66	1.05	–0.32	0.32	0.06	0.23
$\sigma_{AE}^2$	20	–1.89	5.93	1.31	4.46	1.24	5.80	–2.51	10.47	1.18	4.00	2.44	8.63
$\sigma_{DE}^2$	0	0.46	0.38	0.00	1.77	14.79	325.49	28.90	1387.17	0.10	0.45	0.81	7.16
$\sigma_{AAE}^2$	20	–0.43	0.29	–0.90	0.88	–1.94	5.51	–1.33	2.48	–0.18	0.14	–0.12	0.12
$\sigma_e^2$	20	–0.01	0.01	–0.04	0.01	0.03	0.01	–0.11	0.02	–0.15	0.03	–0.06	0.01

† Genetic Designs 1 through 6 (GD1–GD6) are defined in Table 1.

## SIMULATION RESULTS

Different sets of input values have been used in our simulation study and similar conclusions resulted. For simplicity, we only reported the influence of the six GDs on the estimation of each variance component using the ADAA model for three cases: (i) all genetic effects, (ii) no  $A \times A$  epistatic effects, and (iii) no D effects. The simulation results (bias and MSE) are summarized in Tables 2–4, respectively.

The bias measures the difference between mean estimate and the preset value. Negative bias indicates underestimation for a parameter and positive bias indicates overestimation for a parameter. Thus the smaller a bias is, the more unbiased the estimate will be. The MSE measures the precision of an estimate. A smaller MSE indicates a more precise estimation. Simulation results showed that the estimated genetic variance components for A effects,  $A \times E$  effects,  $A \times A$  epistatic effects, and  $A \times A \times E$  effects were unbiased for the six designs (Tables 2, 3, and 4). The variance components of D effects and  $D \times E$  effects could be estimated, but with large bias for Designs 3 and 4; however, estimates were almost unbiased for Designs 1, 2, 5, and 6. Designs 5 and 6 gave similar mean estimated variance components with a little larger MSE, as we expected. Designs 3 and 4 would obtain biased estimation for D and  $D \times E$  variances when  $A \times A$  epistatic effects exist.

Comparing the results in Tables 2, 3, and 4, we observed that the presence or absence of D effects or  $A \times A$  effects did not affect the bias for each variance component for GDs 1, 2, 5, and 6, indicating that variance component estimation under these four GDs are robust under different cases. The variance components for A,  $A \times A$ ,  $A \times E$ , and  $A \times A \times E$ , and residual effects were unbiased estimated with D or  $A \times A$  effects absent or present for GDs 3 and 4 (Tables 3 and 4). It indicated that the estimations for variances components of these effects under GDs 3 and 4 were not influenced by the absence or presence of D effects or  $A \times A$  effects. With all genetic effects present, the D variance was generally underestimated for GDs 3 and 4, while  $D \times E$  variance was generally overestimated (Table 2), indicating that GDs 3 and 4 provide biased estimations for D and  $D \times E$  effects when all genetic effects exist. When there was no  $A \times A$  epistatic effect, the variance components for D effects and  $D \times E$  effects can be estimated with slight bias for GD 3, while D variance can be estimated with a slight bias (–4.06) and variance for  $D \times E$  effects with a large bias for GD 4 (Table 3). With no influence of D effects, D variance were estimated unbiasedly but with large MSE for both GDs 3 and 4, while variance for  $D \times E$  effects were overestimated for these two GDs.

In summary, GDs 5 and 6, which were partial, pro-

**Table 5. Seven combinations from a 2-yr cotton data set (Wu et al., 1995).**

Data set	Environment	Entries
1	1992	10P + 20F <sub>1</sub> + 20F <sub>2</sub>
	1993	10P + 20F <sub>1</sub> + 20F <sub>2</sub>
2	1992	10P + 20F <sub>1</sub>
	1993	10P + 20F <sub>2</sub>
3	1992	10P + 20F <sub>2</sub>
	1993	10P + 20F <sub>1</sub>
4	1992	10P + 20F <sub>1</sub>
	1993	10P + 20F <sub>1</sub> + 20F <sub>2</sub>
5	1992	10P + 20F <sub>2</sub>
	1993	10P + 20F <sub>1</sub> + 20F <sub>2</sub>
6	1992	10P + 20F <sub>1</sub> + 20F <sub>2</sub>
	1993	10P + 20F <sub>1</sub>
7	1992	10P + 20F <sub>1</sub> + 20F <sub>2</sub>
	1993	10P + 20F <sub>2</sub>

vided similar and robust estimations for variance components with slightly larger MSE compared with GDs 1 and 2, which are complete, in two environments. This suggested that partial GDs 5 or 6 could be a good choice when not enough seeds (or individuals) of F<sub>1</sub> and F<sub>2</sub>, or F<sub>2</sub> and F<sub>3</sub> are obtained. It also appeared that the GDs did not have great influence on estimation for A (including A × E) variances, and A × A (including A × A × E) variances though the corresponding MSE might vary. However, some partial GDs such as GDs 3 and 4 had great impact on the estimation of D and D × E variances under some situations, especially when D effects exist.

### Actual Data Example

Ten upland cotton lines and their 20 F<sub>1</sub> and F<sub>2</sub> hybrids were planted at Zhejiang Agricultural University in 1992 and 1993 (Wu et al., 1995). The mating design was the North Carolina Design II: five parents as females were crossed with the other five male parents to produce F<sub>1</sub> seeds in 1991 and 1992 at the same university. Note that only 20 F<sub>1</sub> hybrids had enough seeds for planting so the genetic mating design was partial. F<sub>1</sub> seeds were sent to Hainan Province, China, to produce F<sub>2</sub> seeds in the winter of 1991. The experiment was a completely randomized block design with three replicates each year. The trait *lint percentage* (lint weight divided by cotton weight) was used as an example in this study. From this data set, we organized seven data sets (Table 5), where Data Set 1 was equivalent to GD 1 in Table 1, Data Sets 2 and 3 were equivalent to design 3 in Table 1, and Data Sets 4 to 7 were equivalent to GD 5. These data sets were used to estimate the variance components for lint percentage using the ADAA model. The resampling (jackknife) method was applied to calculate the standard error for each parameter by removal of each replicate within each environment. The approximate *t* test (one-tail) was used to detect the significance of each variance component (Miller, 1974). In this example, *df* = 5. The estimated variance components for this trait are summarized in Table 6. The estimated variance for residual was consistent for seven data sets. Additive variance for lint percentage was significantly detected for all data sets. In general, the results from Data Sets 4, 5, 6, and 7 were closer to those from the full and complete Data Set 1 than from Data Sets 2 and 3. The

results showed an agreement with the simulation studies conducted in this study.

### DISCUSSION

The partial GDs evaluated in this study could be caused by different genotypes, generations, or entry numbers in a multiple-environment experiment. Thus, our definition for the partial GD is different from the partial mating GDs (Griffing, 1956). Our complete GD is a data set that has the same genotypes in all environments regardless of genetic mating designs. Data obtained from such partial GDs can be considered as unbalanced data which usually cannot be appropriately analyzed by the ANOVA method or the GLM method, especially when a complicated genetic model like the ADAA model is applied. The mixed linear model approaches can be used to estimate variance components even though genotypes vary across environments. These approaches provide useful information to the researchers even though they have data sets obtained from partial GDs. The ADAA genetic model and one of the mixed linear model approaches, the MINQUE approach, have been proposed for more than two decades (Rao, 1971; Cockerham, 1980); however, the main purpose of this study is to provide some useful information on how to handle various types of genetic data sets when some data points are missing using the MINQUE approach.

A good GD should provide reliable results and save resource(s). Regarding the ADAA genetic model, our simulation results showed that GDs 5 and 6, which were partial, provided similar estimations for variance components with slightly larger MSE compared with the GDs 1 and 2, which were complete in two environments. This suggested that partial GDs 5 or 6 could be a good choice when enough seeds (or progenies) of F<sub>1</sub> and selfed F<sub>2</sub>, or selfed F<sub>2</sub> and F<sub>3</sub>, are difficult to obtain. Partial GD 3 or 4 would have large impact on the estimation of D and D × E variances, yet only little impact on A and A × A variance components; thus, these two GDs are not recommended for a general use. Our results from an actual data example had good agreement with our simulated results. Therefore, this simulation study provided important guidance for conducting a genetic experiment when it is difficult to run a complete GD in practice. Furthermore, our simulation results suggested that the mixed linear model approaches can also be applied to more complicated genetic models such as plant seed models (Zhu and Weir, 1994a, 1994b) with partial GDs.

Even though the mating design in our example was partial (with five missing crosses), we found similar results when carrying out 500 simulations using complete and partial GDs listed in Table 1 (data not presented). Thus, it appeared that the efficiency of estimating variance component in the ADAA model was more related to the GDs (partial or complete) rather than the mating designs. In the present simulation study, we choose 20 as a preset value for A variance and residual variance and 20 or 0 as a preset value for D variance and A × A



**Table 6. Estimated variance components for lint percentage using ADAA models with Data Sets 1 to 7.**

	DS 1†	DS 2	DS 3	DS 4	DS 5	DS 6	DS 7
$\sigma_A^2$	1.79 ± 0.43	3.63 ± 0.56	1.78 ± 0.79	2.48 ± 0.40	2.11 ± 0.41	1.50 ± 0.50	2.25 ± 0.49
$\sigma_D^2$	2.05 ± 1.19	2.87 ± 2.32	0.00 ± 0.00	1.67 ± 0.74	0.00 ± 0.00	3.16 ± 1.63	0.87 ± 2.98
$\sigma_{AA}^2$	0.52 ± 0.08	0.00 ± 0.00	0.78 ± 0.46	0.41 ± 0.17	0.38 ± 0.16	0.05 ± 0.09	0.30 ± 0.17
$\sigma_{AE}^2$	0.00 ± 0.00	0.00 ± 0.00	0.97 ± 0.55	0.56 ± 0.19	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00
$\sigma_{DE}^2$	2.15 ± 1.54	0.00 ± 0.00	7.56 ± 3.13	3.63 ± 1.16	4.19 ± 1.29	0.77 ± 1.55	0.97 ± 1.42
$\sigma_{AAE}^2$	0.00 ± 0.00	0.57 ± 0.15	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00
$\sigma_e^2$	1.85 ± 0.25	1.90 ± 0.24	1.94 ± 0.31	1.74 ± 0.20	1.81 ± 0.27	1.88 ± 0.22	1.96 ± 0.27

† DS = data set.

epistatic variance. We conducted more simulations than what we reported in this manuscript. This included different inputs of  $G \times E$  interaction effects and  $A \times A$  interactions and similar conclusions resulted (data not shown). Research may be conducted for additional investigations with different preset values. In addition, researchers may investigate more complicated cases, such as missing crosses and/or missing plots. Our computer program provides the capability of evaluating the reliability of a specific GD (not mating designs) before analyzing a real data set. It can also be used to evaluate several potential GDs and choose an appropriate one. The simulation program we developed can be obtained from the authors by request.

Theoretically, increasing the number of parents will lead to higher statistical testing powers and more robust estimation. The number of parents in our simulation study was small compared with the other studies (Tang et al., 1993; Wu et al., 1995; McCarty et al., 2004), yet it still provided useful information on how to conduct experiments with various types of GDs when genetic resources such as seed supply are limited.

In this study, we focused on the ADAA genetic model. In some cases, other high-ordered epistatic effects may be important for some quantitative traits. For example,  $A \times D$  and  $D \times D$  can be important genetic effects. On the basis of the mixed linear model approach, the ADAA genetic model can be extended to include these genetic effects; however, it remains unknown what GDs should be used to estimate these genetic variance components appropriately. This needs to be evaluated with another simulation study.

Estimating the standard error for each parameter is important for a significance test. Two methods, jackknife and bootstrap (Efron, 1982; Davison and Hinkley, 1997), can be used to estimate the standard errors. In this study, we used the jackknife method. This can be done by removal of one or several measurements each time; for example, one genotype within each environment, or one replication (or one block) within each environment. The latter is called the group-based jackknifing method. Then, the mean jackknife values could be tested by an approximate  $t$  test. However, we found the number of cells (genotypes in a replication or a block) could be different across environments (see GDs 5 and 6, Table 1; or Data Sets 4–7 in Table 3). The reasons we chose the block-based jackknife approach were (i) it reduced the computational intensity while retaining the maximum number of genotypes in the data set after removal of one replication within each environ-

ment, and (ii) it kept the same degrees of freedom for  $t$  tests.

## REFERENCES

- Cao, G., J. Zhu, C. He, Y. Gao, J. Yan, and P. Wu. 2001. Impact of epistasis and QTL  $\times$  environment interaction on developmental behavior of plant height in rice (*Oryza sativa* L.). *Theor. Appl. Genet.* 103:153–160.
- Cheverud, J.M., and E.J. Routman. 1996. Epistasis as a source of increased additive genetic variance as population bottlenecks. *Evolution* 50:1042–1051.
- Cockerham, C.C. 1980. Random and fixed effects in plant genetics. *Theor. Appl. Genet.* 56:119–131.
- Comstock, R.E., and H.F. Robinson. 1948. The components of genetic variance in populations of biparental progenies and their use in estimating the average degree of dominance. *Biometrics* 4:254–266.
- Comstock, R.E., and H.F. Robinson. 1952. Estimation of average dominance of genes. p. 494–516. *In* J.W. Gowen (ed.) *Heterosis*. Iowa State College Press, Ames.
- Davison, A.C., and D.V. Hinkley. 1997. *Bootstrap methods and their application*. Cambridge Univ. Press, Cambridge, UK.
- Doebley, J., A. Stec, and C. Gustus. 1995. Teosinte branched1 and the origin of maize: Evidence for epistasis and evolution of dominance. *Genetics* 141:333–346.
- Edwards, J.D., and K.R. Lamkey. 2002. Quantitative genetics of inbreeding in a synthetic maize population. *Crop Sci.* 42:1094–1104.
- Efron, B. 1982. *The jackknife, the bootstrap and other resampling plans*. Capital City Press, Montpelier, VT.
- Eshed, Y., and D. Zamir. 1996. Less-than-additive interaction of QTL in tomato. *Genetics* 143:1807–1817.
- Griffing, B. 1956. Concept of general and specific combining ability in relation to diallel crossing systems. *Aust. J. Biol. Sci.* 9:463–493.
- Goodnight, C.J. 1988. Epistasis and the effects of the founder events on the additive genetic variance. *Evolution* 42:441–454.
- Lark, K.G., K. Chase, F. Adler, L.M. Mansur, and J.H. Orf. 1995. Interaction between quantitative trait loci in soybean in which trait variation at one locus is conditional upon a specific allele at another. *Proc. Natl. Acad. Sci. USA* 92:4656–4660.
- Lee, S.H., K.Y. Park, H.S. Lee, E.H. Park, and H.R. Boerma. 2001. Genetic mapping of QTLs conditioning soybean sprout yield and quality. *Theor. Appl. Genet.* 103:702–709.
- Li, Z.K., S.R.M. Priston, W.D. Park, A.H. Paterson, and J.W. Stansel. 1997. Epistasis for three grain yield components in rice (*Oryza sativa* L.). *Genetics* 145:453–465.
- Liao, C.Y., P. Wu, B. Hu, and K.K. Yi. 2001. Effects of genetic background and environment on QTLs and epistasis for rice (*Oryza sativa* L.) panicle number. *Theor. Appl. Genet.* 103:104–111.
- Lynch, M., and B. Walsh. 1998. *Genetics and analysis of quantitative traits*. Sinauer Assoc. Publ., Sunderland, MA.
- McCarty, J.C., Jr., J.N. Jenkins, and J. Wu. 2004. Primitive accession germplasm by cultivar crosses as sources for cotton improvement I: Phenotypic values and variance components. *Crop Sci.* 44:1226–1230.
- Miller, R.G. 1974. The jackknife—A review. *Biometrika* 61:1–15.
- Rao, C.R. 1971. Estimation of variance and covariance components MINQUE theory. *J. Multivar. Anal.* 1:257–275.
- Saha, S., J. Wu, J.N. Jenkins, J.C. McCarty, Jr., D.M. Stelly, R.G. Percy, D.A. Raska, and O.A. Gutiérrez. 2004. Association of agronomic and fiber traits with specific Pima 3-79 chromosomes in a TM-1 Background. *J. Cotton Sci.* 8:162–169.

- Tang, B., J.N. Jenkins, J.C. McCarty, and C.E. Watson. 1993.  $F_2$  hybrids of host plant germplasm and cotton cultivars: I. Heterosis and combining ability for lint yield and yield components. *Crop Sci.* 33:700–705.
- Wu, J. 2003. Genetic variation, conditional analysis, and QTL mapping for agronomic and fiber traits in upland cotton. Ph.D. diss. Mississippi State Univ., Mississippi State, MS.
- Wu, J., J. Zhu, F. Xu, and D. Ji. 1995. Analysis of genetic effect—5 environment interactions for yield traits in upland cotton (Chinese). *Heredita* 17(5):1–4.
- Xu, Z.C., and J. Zhu. 1999. A new approach for predicting heterosis based on an additive, dominance and additive  $\times$  additive model with environment interaction. *J. Hered.* 82:510–517.
- Zhu, J. 1994. General genetic models and new analysis methods for quantitative traits (Chinese). *J. Zhejiang Agric. Univ.* 20:551–559.
- Zhu, J. 1989. Estimation of genetic variance components in the general mixed model. Ph.D. diss., North Carolina State Univ., Raleigh, NC.
- Zhu, J. 1998. Analytical methods for genetic models. Press of China Agriculture, Beijing, China.
- Zhu, J., and B.S. Weir. 1994a. Analysis of cytoplasmic and maternal effects: I. A genetic model for diploid plant seeds and animals. *Theor. Appl. Genet.* 89:153–159.
- Zhu, J., and B.S. Weir. 1994b. Analysis of cytoplasmic and maternal effects: II. A genetic model for triploid endosperms. *Theor. Appl. Genet.* 89:160–166.
- Zuo, Q., J. Zhu, Y. Liu, X. Pan, and J. Zhang. 2000. Methods of quantitative genetic analysis of plant agronomic traits and yield traits for unequal experiment design (Chinese). *Sci. Agric. Sin.* 33(2):1–9.
- Zuo, Q., J. Zhu, Y. Liu, X. Pan, and J. Zhang. 2001. Methods of genetic analysis for triploid endosperm quality traits for unequal experiment design (Chinese). *J. Biomathematics* 16(1):16–22.